

## *Research Article*

# **Research on the Human Dynamics in Mobile Communities Based on Social Identity**

**Qiang Yan, Lianren Wu, and Lanli Yi**

*School of Economics and Management, Beijing University of Posts and Telecommunications,  
Beijing 100876, China*

Correspondence should be addressed to Qiang Yan, [buptyan@gmail.com](mailto:buptyan@gmail.com)

Received 2 August 2011; Revised 28 October 2011; Accepted 3 November 2011

Academic Editor: Zhen Jin

Copyright © 2012 Qiang Yan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Through analyzing the data about the releases, comment, and forwarding of 120,000 microblog messages in a year, this paper finds out that the intervals between information releases and comment follow a power law; besides, the analysis of data in each 24 hours reveals obvious differences between microblogging and website visit, email, instant communication, and the use of mobile phone, reflecting how people use fragments of time via mobile internet technology. The paper points out the significant influence of the user's activity on the intervals of information releases and thus demonstrates a positive correlation between the activity and the power exponent. The paper also points out that user's activity is influenced by social identity in a positive way. The simulation results based on the social identity mechanism fit well with the actual data, which indicates that this mechanism is a reasonable way to explain people's behavior in the mobile Internet.

## **1. Introduction**

In traditional studies, it is usually assumed that people's behaviors are random in time and thus can be simply described as Poisson processes. However, as the ways of data collection and capability of data processing are becoming significantly improved, more and more empirical studies on people's behaviors prove that many of them deviate from the Poisson distribution. Studies in this field mainly fall into two categories, namely, empirical study on the statistical features of people's behaviors in different contexts especially those in the Internet context; and the study on theoretical models interpreting people's behaviors. This paper focuses on features of people's behaviors in the mobile Internet as well as the theoretical model concerned.

As for the empirical study on statistical features of people's behaviors, Barabási researched first the interval distribution of behaviors like sending emails, waiting for reply

and mail communication [1]. The research reveals that the distribution of intervals between sending emails and getting reply is consistent with power law, with the power exponent being 1, but does not agree with the Poisson distribution, due to the coexistence of long periods of inactivity and bursts of occurrence in a short-time period. In the same year, Oliveira and Barabási studied the correspondences via water and land of Einstein and Darwin and found that these correspondences follow a power law despite the difference from email [2]. According to analysis of time recorded in an MMORPG server, Grabowski and Kruszewska pointed out that the possibility of individuals spending time on an activity follows a power law [3]. Gonçalves and Ramasco found that the distribution of intervals between the visits of a specific user to the same website also follows a power law, with the exponent being 1.25 [4]. Vojnović conducted a statistic analysis of a search log of mobile network over one week supplied by a US mobile service provider and revealed that the distribution of intervals between search via mobile devices in a day follows a power law [5]. Chmiel et al. pointed out that the distribution of time spent on browsing subpages of portal websites follows a power law [6]. Based on analysis of people's behaviors, Yu and other scholars pointed out the numbers of browsing and replies in online forums follow a power law, with the exponent between 3.5 and 1.6, and times of checking and number of replies follow nonlinear power distribution [7]. Hu and Han studied a log data of online music system and found that the distribution of intervals between songs consecutively ordered has a heavy tail, and the heavy-tail distribution becomes more and more irregular as the users' activities are reduced [8]. Jiang et al. analyzed the number of roles in an online role-playing game and found that the fluctuation of the user number follows a non-Gaussian distribution with a peak and a tail following a power law [9]. Candia et al. provided cases of heavy-tail distribution of intervals between consecutive calls by mobile phone [10]. Wu et al. analyzed messaging between two users, which revealed that the distribution of intervals between messages is not entirely a Poisson distribution or follows a power-law distribution, but a bimodal distribution combining the two [11]. According to research on movie ordering, the interval distribution follows a power law, with the exponent being 2.08, and the power exponent is monotonically correlated to the activity of the corresponding filmgoers [12]. Strongly positive correlations between the individual activity and power law exponent of intervals distribution have also been proved in [13–15].

As for models explaining people's behaviors, Barabási and Vázquez and others pointed out that the bursty nature of human behavior is a consequence of a decision-based queuing process: when individuals execute tasks based on some perceived priority, the timing of the tasks will be heavy tailed, most tasks being rapidly executed, with a few experiencing very long waiting times [1]. According to Vázquez et al., the response time distribution of the tasks could in fact drive the interval distribution and both the two distributions should decay with the same exponent [16]. This has also been proved by other researches [1, 17]. Blanchard and Hongler proposed an "aging mechanism" which connects the priority of each task with time and assumed the "Earliest Deadline First" principle [18]. Based on Barabási's research, Deng and other researchers added the task deadline as a restrictive condition and analyzed the influence of the deadline on the waiting time of the task [19]. Dall'Asta et al. interpreted the process of task fulfillment from the perspective of economic optimum [20]. Vázquez pointed out that people perceive their past activity ratio and may increase or reduce their activities accordingly [21]. In other words, memory, an important attribute of human beings, significantly influences the features of human dynamics. Through statistical analysis of browsing times and number of replies, Yu et al. concluded that people reply in BBS according to their interests and are hardly influenced by

their memories [7]. Early replays cast no impact on later visitors, and people just reply as they like.

People behave because of complex motivations, and not all things are tasks, for example, browsing and ordering movies, but such behaviors follow certain statistical laws too. According to Shang and other scholars, people's interest in new things wanes or even disappears due to frequent involvement, but may suddenly revive after lasting indifference. The change of people's interest may cause the heavy-tail distribution of their behaviors. Shang and others proposed to quantify people's interest by the probability of occurrence during intervals [22]. Han and others also noticed the fact that people's interest in a certain activity may be changed due to their feelings, and thus proposed the self-adapting human dynamic mechanism [23]. Guo held that blogging is becoming less attractive as people are involved more and more [24]. Wu and others established a human dynamic model, holding that individuals' behaviors in online commenting system are influenced by other individuals [25].

Existing studies have proved that the Internet has laid a sound data foundation for human dynamic research, and its wide use in people's daily life also determines the significance of Internet-based human dynamics research. However, the above-mentioned research results are challenged with the rapid development of the mobile Internet in recent years.

In the context of a mobile Internet, people's behaviors in fragments of time are hardly "task oriented" as they perform according to what they see and feel. This shows that the priority-based task queue model is limited when explaining such behaviors. A dynamic model with memory, as well as an interest-driven dynamic model, is also limited due to the fragmentation of time and content for behaviors in the mobile Internet. Moreover, most users of the mobile Internet are youngsters who try to express their unique personality on the one hand and on the other hand desire a social identity which is manifested in behaviors like online comment, collecting, ordering, and forwarding. Therefore, the social identity as a driving force might supplement interest-driven human dynamics research.

This paper analyzes data about microblog on <http://www.sina.com/>, which is among the largest Chinese websites for microblog in the world. Empirical analysis shows that the distribution of intervals between information releasing on microblog follows a power law and verifies a positive correlation between user's activity and the power exponent. Based on rescaling method, our research presents a universal behavior for users with different activities, which is in consistency with the results in [10, 13, 15]. The paper assumes that the activity of individual user is influenced by social identity and quantifies the social identity to improve the memory model [5]. The simulation results are in consistency with the empirical data.

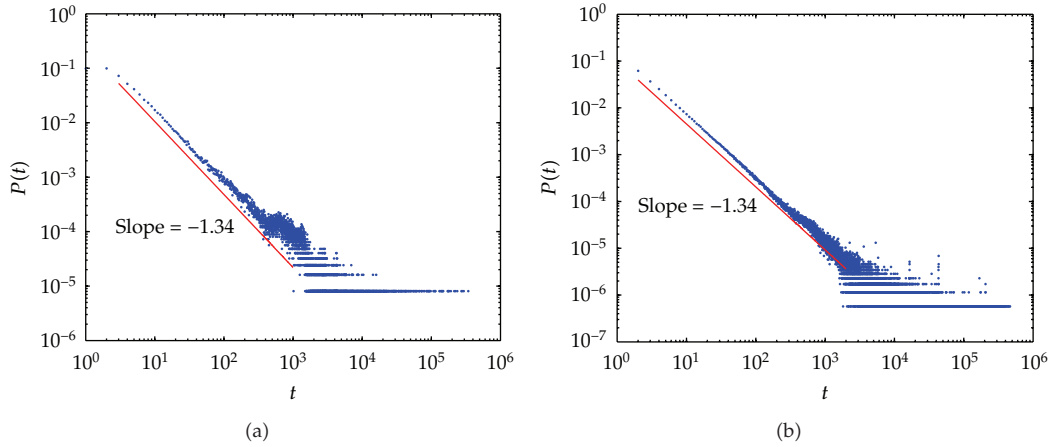
The remainder of the paper is organized as follows. Section 2 presents the empirical analysis results. Section 3 analyzes the influence of user's activity on statistical features of people's behaviors, proposes a social identity mechanism, and compares the simulation-based results with empirical data. Finally, the conclusions of the whole paper are presented in Section 4.

## 2. Statistic Features of Empirical Data

This research collected microblog messages from <http://www.sina.com/>. The website organizes messages by different topics. Considering that topics about entertainment have

**Table 1:** An example of message releasing behaviors.

User ID	Message ID	Releasing time
u1	m1	t1
u2	m2	t2
u1	m3	t3
u1	m4	t4
u2	m5	t5
u2	m6	t6
—	—	—

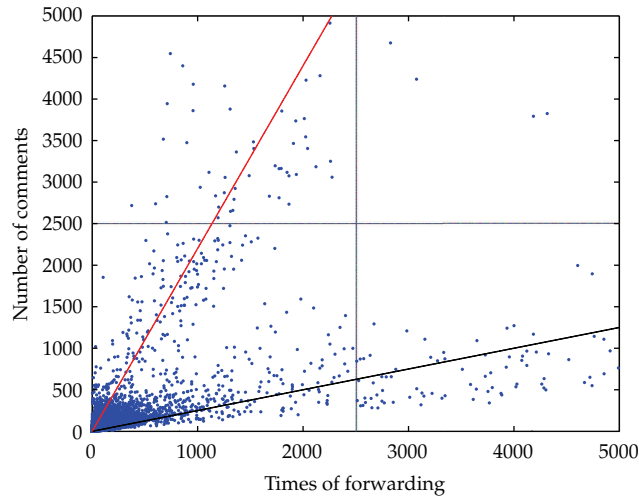
**Figure 1:** Distribution of intervals (a) distribution of intervals between initial releases (b) distribution of intervals between comments.

aggregated many users at current stage, we collected all messages about an entertainment topic, ranging from August 20, 2009 to September 3, 2010. During this period, there are totally 125,152 messages released by 175 users, which have been forwarded 2,260,826 times and triggered 1,786,000 comments.

To analyze the releasing behavior, releasing intervals of each individual are calculated separately. Then intervals of all individuals are aggregated to figure out the distribution. For example, suppose Table 1 is the releasing information of all individuals. Each line of the table presents the releasing time of every message from corresponding users. For user u1, releasing intervals are  $(t_3-t_1)$ ,  $(t_4-t_3)$ , and so on. For user u2, releasing intervals are  $(t_5-t_2)$ ,  $(t_6-t_5)$ , and so on. Intervals of comments are calculated similarly.

According to the data collected, the interval distributions of both releasing and commenting follow power law in the log-log grid, as indicated in Figure 1.

According to Figure 1, the power-law exponents of message releasing and commenting intervals are almost the same, which is consistent with the results of [16, 17]. Individual's message releasing behavior is driven by commenting behavior of others to some extent. Thus both the two distributions decay with the same exponent. However, the exponent 1.34 in this paper is higher than that in the research about email communication [16], and lower than the exponent concluded in [17]. In email communication, spam and irrelevant emails often mix with normal emails, which may cause the long tail of the interval distribution. However, in the one hand, microblog messages in <http://www.sina.com/> are organized by topics, and



**Figure 2:** Correlation between comments and forwarding.

communications in microblog communities are in a way of multipoint to multipoint, which improves the frequency of communications. In the other hand, microblog is based on mobile Internet technologies that enable individuals to release messages at anytime and anywhere. The differences in the way of communication and technologies may cause the diversity of the power law exponents in different scenarios.

Figure 2 explains the correlation between the number of comments and the times of forwarding.  $x$  refers, the times of forwarding to and  $y$  means the number of comments. When  $x$  is lower than 2,500,  $y = 2.2x$ . When  $y$  is lower than 2,500,  $y = 0.25x$ . There is almost no information in the area where both  $x$  and  $y$  are above 2,500. In over 95% of all cases, both  $x$  and  $y$  are lower than 500, and the two are not obviously correlated. On this basis, messages on microblogs can be divided into two groups: those more frequently forwarded while less commented, which are intended to inform more individuals rather than attract comments, and those more commented but less forwarded, which may arouse extensive discussion, instead of wide spread.

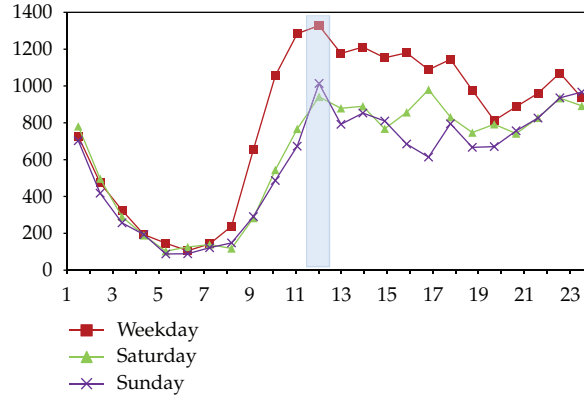
As indicated in Figure 3 about the analysis of microblog data over 24 hours, information releases peak between 11 am and 12 am both at weekends and on weekdays, and that is quite different from the peak of using email, instant communication devices, and mobile phones which is usually at 10 am. This lagging peak of microblog indicates that most messages on microblogs have nothing to do with work and are just small talks in break time.

Figure 4 illustrates the releasing time of all messages from the 175 users in one day. Two peak time periods, namely, 11:00–13:00 and 23:00–01:00, again reflect the features of using microblog in fragments of time.

### 3. Social Identity-Based Human Dynamic Model

#### 3.1. Influence of User's Activity on Statistical Features

The intensity of messaging (activity) has significant influence on distribution of intervals of messaging, and the more active the user is, the larger the power exponent for interval



**Figure 3:** Microblog data over a 24 h period. Note. The figure is about the number of messages posted on microblog in 24 hours. “1” on the X-axis indicates from “0” to “1” o’clock, “2” means from “1” to “2”, and so on. The weekday curve represents the average number between Monday and Friday. The Y-axis is the number of messages.



**Figure 4:** The releasing time of messages sent by all 175 users in one day.

distribution will be. Suppose the activity of user  $i$  is  $A_i = n_i/T_i$ ,  $n_i$  indicating the total sum of messages sent by the user,  $T_i$  being the total time spent on by all behaviors. In this paper,  $A_i$  means the number of messages sent by user  $i$  per day. To explore the role of user’s activity in the online social commenting system, users sending over 1,000 messages are picked out. As shown in Figure 5, the average activity of these users is  $\langle A \rangle = 12.04$ .

These users are arranged in descending order of their activity and then divided into 5 groups with each including the same number of members. As for the average activity of each group,

$$\langle A \rangle_1 > \langle A \rangle_2 > \langle A \rangle_3 > \langle A \rangle_4 > \langle A \rangle_5. \quad (3.1)$$

Figure 6 illustrates the interval distributions of the five groups. According to the figure, the decay exponents depend on the activity of the users, and therefore the interval distributions corresponding to different levels of activity are more representative than a global one. Lower average activity corresponds to lower power exponent and thus allows longer intervals, which is in consistency with the conclusion in [12]. The correlation between exponents and corresponding average activities is presented in Figure 7.

The variation of activity reflects the difference of individuals’ behavior patterns. Here we use the recalling method in order to describe the universal behavior of different groups. Instead of considering the values for interval  $t$ , we take into account the rescaled variable  $t/\Delta t$ ;  $\Delta t$  represents the average interval of the respective group of users. As illustrated in Figure 8, the scaling produces a data collapse between the different curves of five groups. This phenomenon has also been noticed in other systems [10, 13, 15].

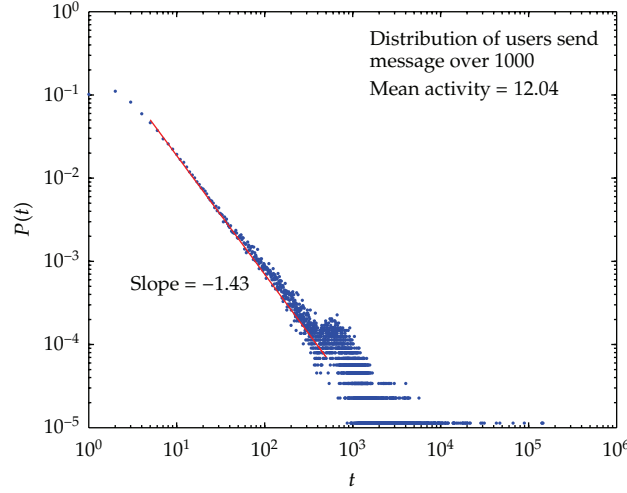


Figure 5: Intervals of messaging for all users sending over 1,000 messages.

### 3.2. Influencing Mechanism of Social Identity on Information Releases Behaviors

To analyze the influences on user's activity in mobile communities, this paper introduces social identity which is quantified and used to improve the memory model proposed by Vázquez. Based on intuitive experience on users' messaging habits, if messages from user  $i$  are recognized by other users, namely, if such messages are commented or forwarded, user  $i$  would like to send more messages, otherwise he may become less interested. According to the equation proposed in reference [21],

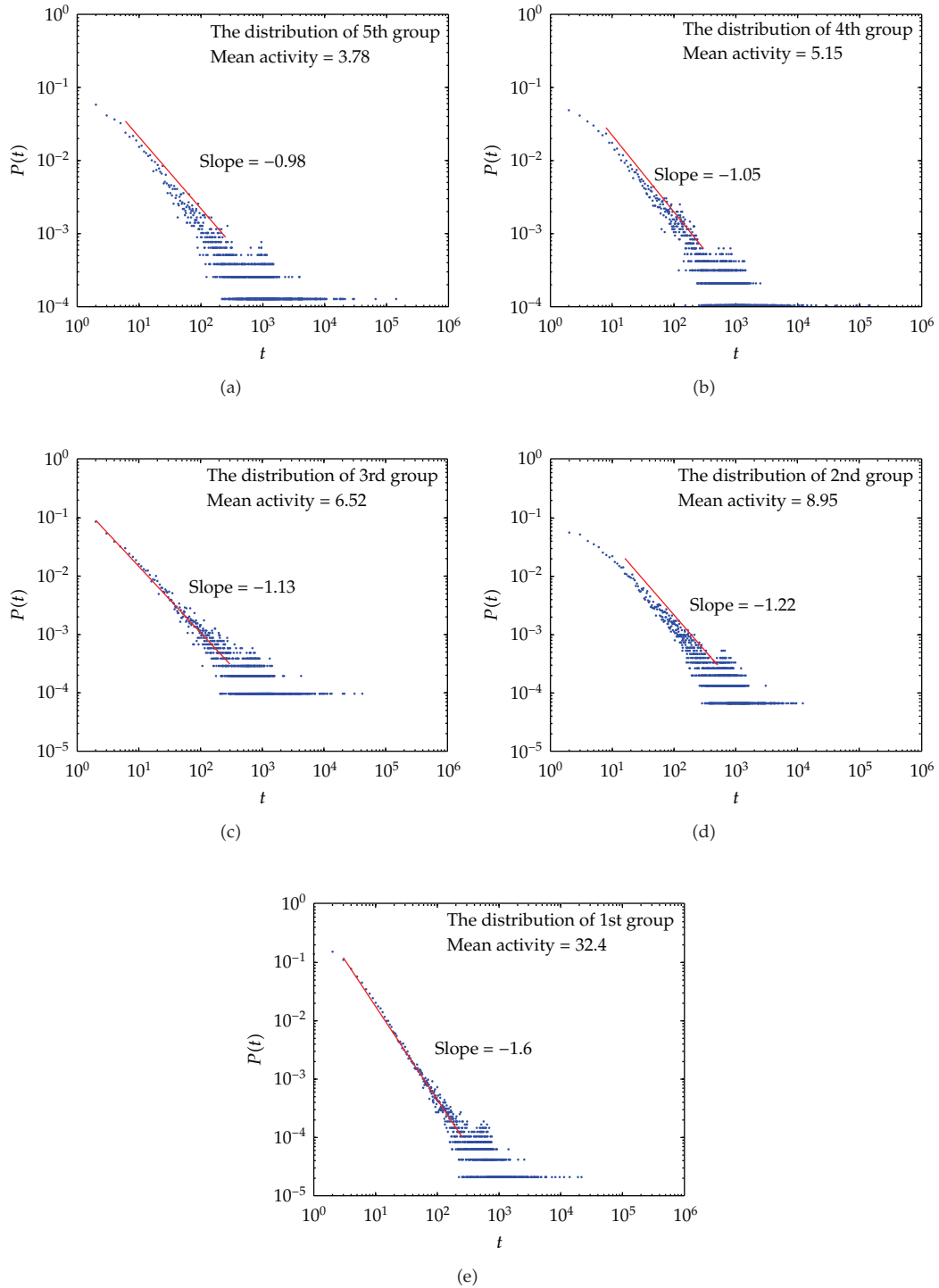
$$\lambda(t) = a \frac{1}{t} \int_0^t \lambda(s) ds, \quad (3.2)$$

the parameter  $a$  controls the degree and type of intuitive perception. In this paper,

$$a = \begin{cases} \frac{1}{(1-p)} & 0 < p < 1, \\ 1-p & \end{cases} \quad (3.3)$$

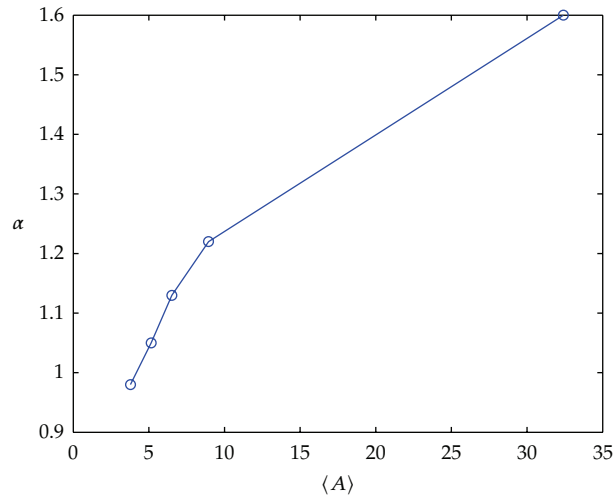
$p$  means the probability of sending another message after the user's previous message is commented or forwarded. In case of no comment or forwarding, the possibility of sending another message is  $1-p$ . The value of  $p$  approximates the ratio between the number of messages commented or forwarded and the total number of messages sent by the user. From (3.2) we can get the probability density function following power law:

$$f(\tau) = \frac{1}{\tau_0} \frac{a}{a-1} \Gamma\left(\frac{1-2a}{1-a}\right) \left(\frac{\tau}{\tau_0}\right)^{-a}. \quad (3.4)$$

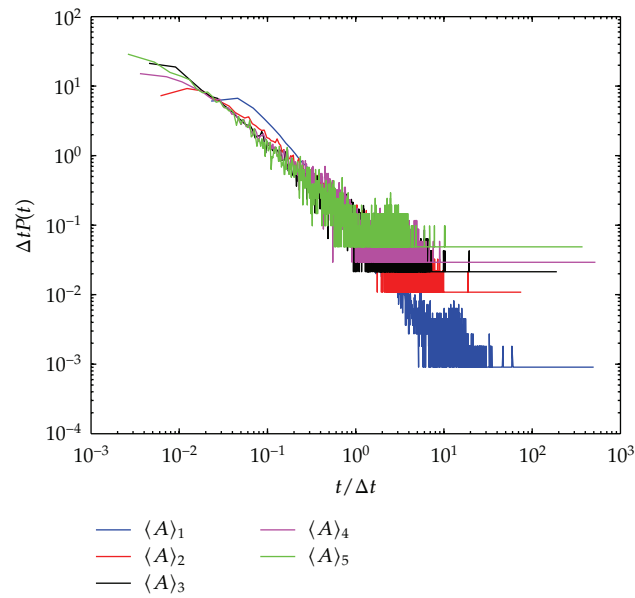


**Figure 6:** Interval distributions of users with different activity. Subgraph (a) to (e) refers to interval distribution for group 1 to 5, respectively.





**Figure 7:** Correlation between exponents and corresponding average activities.



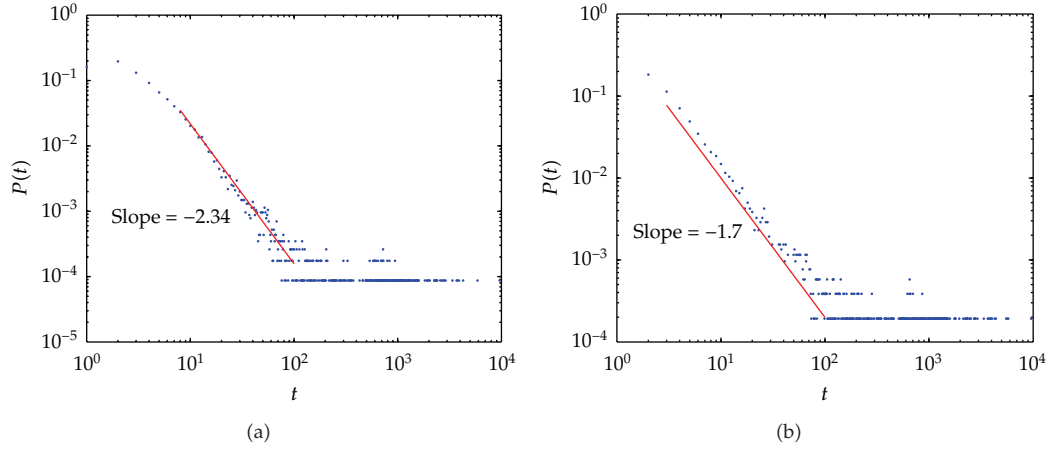
**Figure 8:** The interval distributions of every group.

In case of comment or forwarding, the social identity mechanism promotes users to send more messages ( $a > 1$ ), and the corresponding power exponent is

$$\alpha = 2 + \frac{1}{a-1}. \tag{3.5}$$

**Table 2:** Messaging data of user 2 and user 4.

User number	Total messages sent	Number of messages commented or forwarded	Number of messages receiving no comment or not forwarded	$p$
2	13960	10114	3846	0.72
4	1619	1295	324	0.80

**Figure 9:** Distribution of intervals between messages sent by user 2 under the social identity mechanism.

If no comment or forwarding occurs, the social identity is discouraging ( $a < 1$ ), and the corresponding power exponent is

$$\alpha = 2 - \frac{a}{1-a}. \quad (3.6)$$

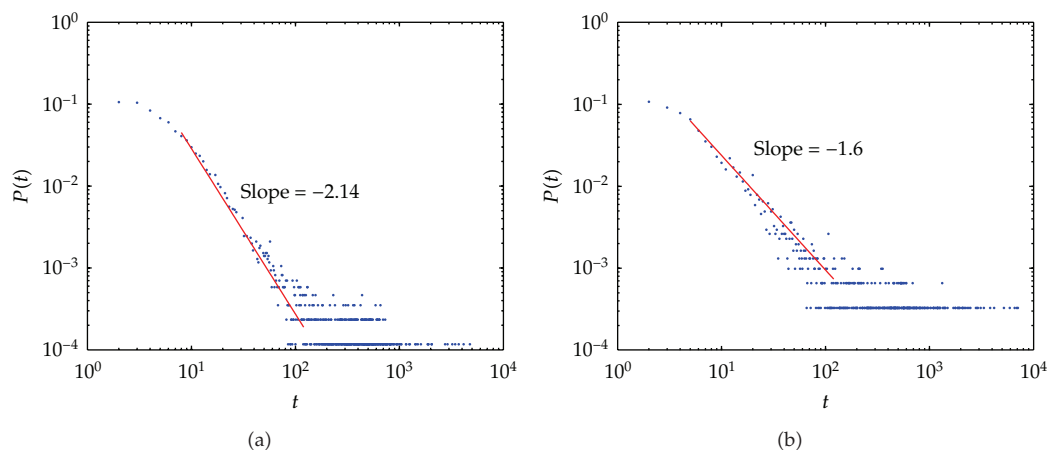
### 3.3. Comparison between Simulation Results and Empirical Data

Two typical users are picked out, respectively, from group 2 and group 4 and are marked as user 2 and user 4. Their information is as shown in Table 2.

The social identity mechanism accelerates message sending in case of  $0.7 < p < 0.9$ , then the range of corresponding power exponent obtained via formula (3.5) is  $2.11 < \alpha < 2.43$ . While the social identity mechanism is discouraging, the range of corresponding power exponent calculated according to the same formula is  $1.57 < \alpha < 1.89$ .

Figure 9 is about the distribution of intervals between messages sent by user 2 under the social identity mechanism. Figure 9(a) tells the distribution of intervals between messages sent by user 2 in case of comment or forwarding, with  $p = 0.72$  and  $\alpha = 2.34$ . Figure 9(b) shows the distribution of intervals between messages in case of no comment or forwarding, with  $1 - p = 0.28$  and  $\alpha = 1.70$ . The values of power exponents are in ranges calculated when the social identity mechanism is encouraging or discouraging.

Figure 10 deals with the distribution of intervals between messages sent by user 4 under the social identity mechanism. Figure 10(a) tells the distribution of intervals between messages sent by user 4 in case of comment or forwarding, with  $p = 0.8$  and  $\alpha = 2.14$ .



**Figure 10:** Distribution of intervals between messages sent by user 4 under the social identity mechanism.

Figure 10(b) shows the distribution of intervals between messages in case of no comment and forwarding, with  $1 - p = 0.20$  and  $\alpha = 1.60$ . The values of power exponents are in ranges calculated when the social identity mechanism is encouraging or discouraging.

The model-based results fit well with the empirical data, indicating the influence of social identity on the activity of individual users. Namely, the higher the social identity is, the more active the user will be, and vice versa.

## 4. Conclusions

In recent years, many studies have been conducted on the statistical law of people's behavior in the context of the Internet and theoretical mechanisms concerned and found that many behaviors follow power-law distribution in terms of statistical features about time. This paper analyzes microblog and reveals the law of information releases in mobile communities.

Analysis of over 120,000 messages sent through microblog at <http://www.sina.com/> reveals a power-law interval distribution; besides, the law of microblogging differs from that of email, instant communication, and mobile phone calls, according to the 24-hour behaviors analysis. The mobile internet technology-based microblog enables users to send messages by mobile phone whenever and wherever they like, rather than only via computer. This paper explores the features of using mobile Internet technology in fragments of time and the difference between that and online browsing in terms of time and space. In addition, it points out the significant influence of user's activity on distribution of intervals between information releases and testifies the positive correlation between the user's activity and the power exponent. What is more important is that it finds that the social identity has direct influence on the user's activity in mobile communities: the higher the social identity is, the more active the user will be, and vice versa. Results from social identity-based simulation are in consistency with actual data, indicating that the social identity mechanism is a proper way to interpret people's behaviors in the context of the mobile internet.

## Acknowledgments

This work was supported by Program for New Century Excellent Talents in University, NSFC (70671010), Key Project of Chinese Ministry of Education. (108011), Joint Construction Project of Beijing Municipal Commission of Education (XK100130439) and the Fundamental Research Funds for the Central Universities (2009RC1005).

## References

- [1] A. L. Barabási, "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, no. 7039, pp. 207–211, 2005.
- [2] J. G. Oliveira and A. L. Barabási, "Darwin and Einstein correspondence patterns," *Nature*, vol. 437, no. 7063, p. 1251, 2005.
- [3] A. Grabowski and N. Kruszewska, "Experimental study of the structure of a social network and human dynamics in a virtual society," *International Journal of Modern Physics C*, vol. 18, no. 10, pp. 1527–1535, 2007.
- [4] B. Gonçalves and J. J. Ramasco, "Human dynamics revealed through Web analytics," *Physical Review E*, vol. 78, no. 2, Article ID 026123, 7 pages, 2008.
- [5] M. Vojnović, "On mobile user behavior patterns," in *Proceedings of the International Zurich Seminar on Communications*, pp. 26–29, Zurich, Switzerland, 2008.
- [6] A. Chmiel, K. Kowalska, and A. H. Janusz, "Scaling of human behavior during portal browsing," *Physical Review E*, vol. 80, no. 6, Article ID 066122, 7 pages, 2009.
- [7] J. Yu, Y. Hu, M. Yu, and Z.-R. Di, "Analyzing netizens' view and reply behaviors on the forum," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 16, pp. 3267–3273, 2010.
- [8] H.-B. Hu and D.-Y. Han, "Empirical analysis of individual popularity and activity on an online music service system," *Physica A*, vol. 387, no. 23, pp. 5916–5921, 2008.
- [9] Z.-Q. Jiang, W.-X. Zhou, and Q.-Z. Tan, "Online-offline activities and game-playing behaviors of avatars in a massive multiplayer online role-playing game," *Europhysics Letters*, vol. 88, no. 4, Article ID 48007, 2009.
- [10] J. Candia, M. C. Gonzalez, P. Wang, T. Schoenharl, G. Madey, and A. L. Barabasi, "Uncovering individual and collective human dynamics from mobile phone records," *Journal of Physics A: Mathematical and Theoretical*, vol. 41, no. 22, Article ID 224015, 2008.
- [11] Y. Wu, C.-S. Zhou, J.-H. Xiao, J. Kurths, and H.-J. Schellnhuber, "Evidence for a bimodal distribution in human communication," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 44, pp. 18803–18808, 2010.
- [12] T. Zhou, H. A. T. Kiet, B. J. Kim, B. H. Wang, and P. Holme, "Role of activity in human dynamics," *Europhysics Letters*, vol. 82, no. 2, Article ID 28002, 2008.
- [13] F. Radicchi, "Human activity in the web," *Physical Review E*, vol. 80, no. 2, Article ID 026118, 2009.
- [14] Z.-D. Zhao, H. Xia, M.-S. Shang, and T. Zhou, "Empirical analysis on the human dynamics of a large-scale short message communication system," *Chinese Physics Letters*, vol. 28, no. 6, Article ID 068901, 2011.
- [15] P. Wang, X.-Y. Xie, C.-H. Yeung, and B.-H. Wang, "Heterogenous scaling in interevent time of on-line bookmarking," *Physica A*, vol. 390, no. 12, pp. 2395–2400, 2011.
- [16] A. Vazquez, J. G. Oliveira, Z. Dezsö, K. I. Goh, I. Kondor, and A. L. Barabási, "Modeling bursts and heavy tails in human dynamics," *Physical Review E*, vol. 73, no. 3, Article ID 036127, pp. 1–19, 2006.
- [17] N.-N. Li, N. Zhang, and T. Zhou, "Empirical analysis on temporal statistics of human correspondence patterns," *Physica A*, vol. 387, no. 25, pp. 6391–6394, 2008.
- [18] P.-H. Blanchard and M. O. Hongler, "Modeling human activity in the spirit of Barabasis queuing systems," *Physical Review E*, vol. 75, no. 2, Article ID 026102, 8 pages, 2007.
- [19] Z.-D. Deng, N. Zhang, and J.-M. Li, "Influence of deadline on human dynamic model," in *Dynamic Model of Human Behavior*, J. L. Guo, T. Zhou, N. Zhang, and J. M. Li, Eds., pp. 29–34, Shanghai System Science Publishing House, Hong Kong, 2008.
- [20] L. Dall'Asta, M. Marsili, and P. Pin, "Optimization in task-completion networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 2, Article ID P02003, 2008.
- [21] A. Vázquez, "Impact of memory on human dynamics," *Physica A*, vol. 373, pp. 747–752, 2007.

- [22] M.-S. Shang, G.-X. Chen, S.-X. Dai, B. H. Wang, and T. Zhou, "Interest-driven model for human dynamics," *Chinese Physics Letters*, vol. 27, no. 4, Article ID 048701, 2010.
- [23] X.-P. Han, T. Zhou, and B.-H. Wang, "Modeling human dynamics with adaptive interest," *New Journal of Physics*, vol. 10, Article ID 073010, 2008.
- [24] J.-L. Guo, "Empirical study and modeling of human behavior dynamics of comments on Blog posts," *Journal of Application Research of Computers*, vol. 28, no. 4, pp. 1422–1433, 2011.
- [25] Y. Wu, C.-S. Zhou, M.-Y. Chen, J. Xiao, and J. Kurths, "Human comment dynamics in on-line social systems," *Physica A*, vol. 389, no. 24, pp. 5832–5837, 2010.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

